DOCUMENT RESUME

ED 078 046 TM 002 861

AUTHOR	Cooper, Mildred P., Ed.
TITLE	Special Report on Criterion-Referenced Test Development, Mid-Atlantic Region Interstate Project 1972-73.
INSTITUTION	West Virginia State Dept. of Education, Charleston.
PUB DATE	[73]
NOTE	42p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Criterion Referenced Tests; Evaluation Techniques; Measurement Instruments; *Norm Referenced Tests; Standardized Tests; Technical Reports; *Test Construction; *Testing

ABSTRACT

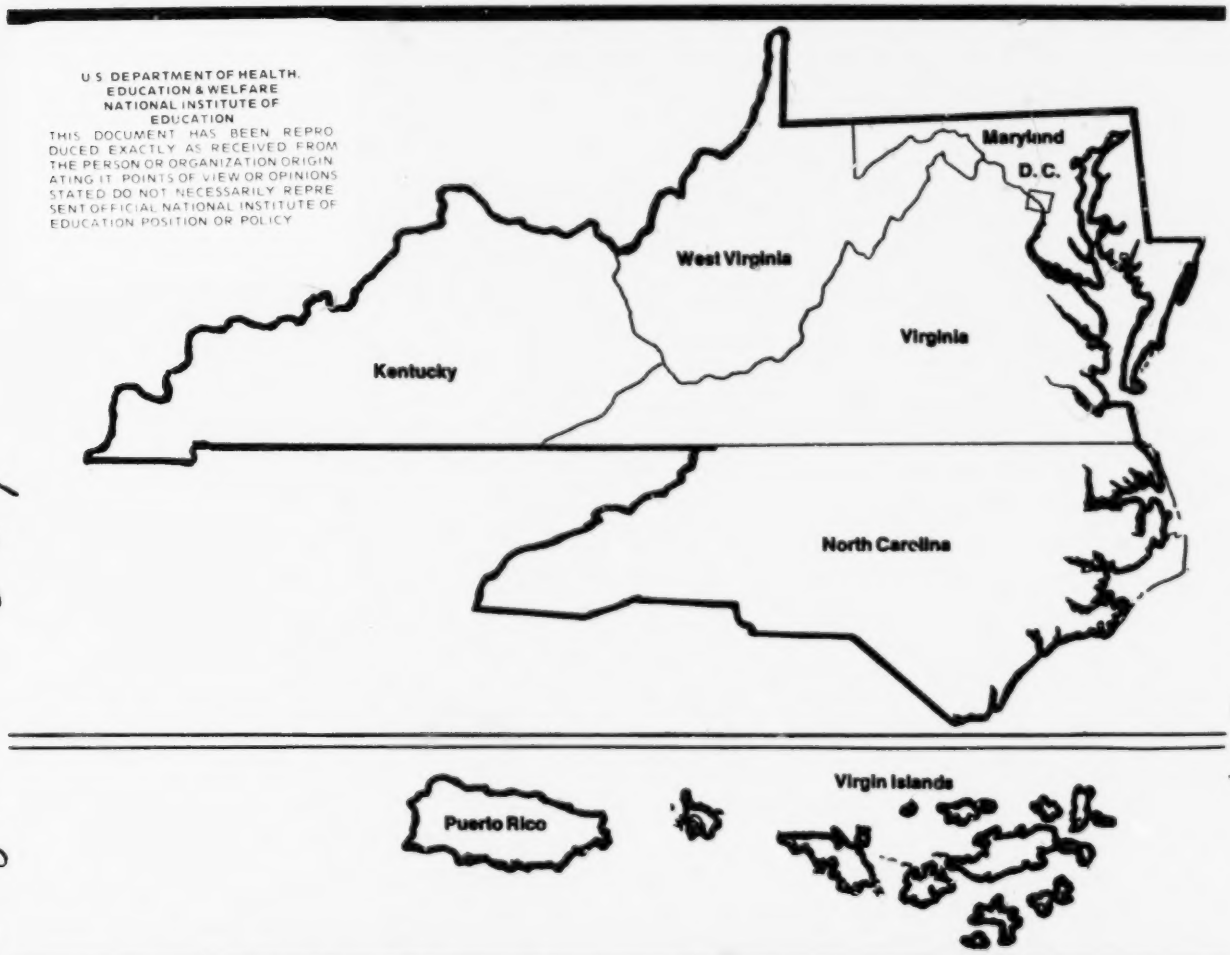
A collection is presented containing four papers dealing with the subject of criterion-referenced testing. This first paper deals with criterion-referenced tests utilizing norm-referenced test items. Aspects covered include a defense of norm-referenced and criterion referenced tests and a justification for the adaptation of norm-referenced test items as a criterion-reference measure. Finally, it is pointed out that the most attractive factors involved in adapting existing norm-referenced test items to criterion-referenced tests are the time- and money-saving possibilities. (For three related papers, see TM 002 862-864) (CK)

FILMED FROM BEST AVAILABLE COPY

ED 078046

SPECIAL REPORT ON CRITERION-REFERENCED TEST DEVELOPMENT

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY



TM 002 861 - 864

Mid-Atlantic Region Interstate Project
1972 - 73

PARTICIPATING STATES: District of Columbia, Kentucky, Maryland, North Carolina, Puerto Rico, Virgin Islands, Virginia and West Virginia

ED 078046

Special Report On Criterion-Referenced Test Development

Mid-Atlantic Region Interstate Project 1972-73

Edited By
Mildred P. Cooper
Public Schools of the District of Columbia

Financed by Funds Provided Under The Elementary and Secondary Education Act of 1965 (Public Law 89-10, Title V, Sec. 505) and sponsoring States.

PARTICIPATING STATES: District of Columbia, Kentucky, Maryland, North Carolina, Puerto Rico, Virgin Islands, Virginia and West Virginia (administering state)

published by:
West Virginia Department of Education
Charleston, West Virginia

PREFACE

Criterion-referenced testing is the newest and fastest growing accountability technique today in our elementary and secondary schools across the nation. States, large city school systems, counties and small local school districts all are interested in utilizing criterion-referenced testing in their assessment programs. Because of this trend, the representatives of the member states of the Mid-Atlantic Region Interstate Planning Project requested that a part of the December, 1972 meeting of the project be allocated to a presentation of papers related to criterion-referenced testing. It was decided that individuals on the staffs of project member states were, as a result of their experiences, well qualified on certain aspects of criterion-referenced measurement. Therefore, the program was planned to include reports from four member states and to provide ample opportunity for discussion.

The frame of reference for all presentations was the definition of criterion-referenced tests given by Robert Glaser in the 1971 edition of Educational Measurement, "a criterion-referenced test is one that is deliberately constructed to yield measurements that are directly interpretable in terms of specific performance standards." The scope of the papers was from the rationale for criterion-referenced testing to comparisons with standardized norm-referenced tests to development plans and activities to recommendations related to criterion-referenced testing.

The collection which is presented here is offered to educators across the country with the hope that the experiences of the Mid-Atlantic Interstate Project member states, their school systems and their staffs, will be of assistance to other educators in their work with assessment programs.

Mildred Pivetz Cooper

CONTENTS

1. Criterion-Referenced Testing and the Adaptation of Norm-Referenced Test Items As a Criterion Reference Measure

David Shannon, Kentucky State Department of Education 1

2. Criterion-Referenced Test Development - A Contractual Arrangement Between the Public Schools of the District of Columbia and a Commerical Test Publisher 9

Mildred P. Cooper, Public Schools of the District of Columbia

3. The Development of Criterion-Referenced Tests in Reading 20

Richard M. Petre, Maryland State Department of Education

4. The Development of a Criterion-Referenced Test of Mathematics 31

Fanny Freytes, Department of Education, Puerto Rico

Criterion-Referenced Testing and the Adaptation of Norm-Referenced Test Items As a Criterion Reference Measure

David Shannon
Director, Office of Planning and Research
Kentucky State Department of Education

INTRODUCTION AND BACKGROUND

Accountability is an old concept in American education that is fathering a new generation of educational processes as still more strategies are aimed at supplying adequate and meaningful accountability models. Cost effectiveness has long been the factor which has provided the thrust for fiscal and management accountability efforts. Since the 1960's, additional thrust has been supplied through the concepts of performance accountability which has grown with the advent of the federal programs which have had evaluation measures built in. Testing programs have received a considerable impetus as a result of these program requirements.

Standardized tests have been widely used and just as widely interpreted as school districts have attempted to evaluate pupil progress and program effectiveness through results of standardized norm-referenced tests. For example, Kentucky has moved from twenty-six (26) pilot testing programs testing 11,447 pupils in the school year 1965-66 to one hundred and fifty-six (156) testing programs in LEA's testing 200,000 pupils in 1970-71 school year. The Elementary and Secondary Education Act, Title I requirements have further encouraged the use of sub-tests

of the standardized tests to evaluate specific goals.

Standardized norm-referenced achievement tests have become widely used throughout the country in testing programs designed to measure student performance as a criterion for accountability. From a psychometric point of view, much has been written about the dangers of using standardized norm-referenced test scores to evaluate the progress of individuals.

Robert Glaser was among the first to move in the promising direction of utilizing test items that are derived directly from a specific well defined objective or performance standard. He pointed out the distinction between this type of test called "criterion-referenced tests" which measures what the pupil can do and norm-referenced tests which compares a pupil's progress with that of others (Lipe and Jung 1971, RER, October 1971). Thus, the intended use of the test results determines the type of test which will be most appropriate. Norm-referenced tests can provide normative data to compare school systems within the nation, to evaluate programs within a system and to a lesser extent to indicate pupil achievement ranking within a grade or a classroom. Criterion-referenced tests offer a more specific and valid means of assessing individual performance and the accomplishment of specific program goals.

The current trend in education is a preciseness in the determination of program goals in terms of behavioral objectives and measurable operational terms. This has created new problems in evaluation. The norm-referenced tests failed to adequately test these goals which were different for each program. Thus, the selection of adequate and appropriate custom-designed measurement devices which could measure attainment of specific objectives became a pressing need. Many State Departments of Education have faced assessment problems as a result of the time constraints, inadequate budgets, and the speed with which many assessments have been mandated. The standardized testing resources at hand have been utilized with varying degrees of effectiveness in meeting assessment needs. Adaptation of reliable and accepted testing materials has resulted. Adapting test items to measure specific program outcomes, a new concept in utilizing existing standardized testing materials, has resulted in criterion referenced tests that adequately meet accountability needs.

Defense of Norm-Referenced and Criterion Referenced Tests

A. V. Nitko has concluded in his definitions of criterion referenced tests that, "in short it is the use to which test results are put that determines their nature and the construction methodology." In instruction, various procedures cannot be considered independently of the instructional context in which they will be used. Particularly important is the integration of test design with instructional design. This consideration has led educators to develop unique criterion referenced tests that focus on an individual's performance attainment in mastering program objectives.

Standardized norm-referenced tests have earned a respected position in educational achievement testing programs when their intended uses have been carefully observed. The expertise of those who construct and revise the major published standardized tests is beyond that of any organization whose sole purpose is other than test construction. Despite the disillusionment experienced by many educators with standardized norm-referenced tests, the facts support their use when they are appropriately selected for the intended use, administered and interpreted correctly, and when the results are reported adequately and understandably. There has thus far been no acceptable alternative offered by the critics to accomplish the intended uses of standardized norm-referenced tests. These uses are specifically to provide normative data to compare school systems within the nation or lesser geographic regions, to provide bench marks of pupil achievement in broad general knowledge bases when compared to those of other pupils, individually or in groups. Such bench marks of comparative rankings can point out weaknesses in programs which may be meeting local needs yet be ignoring the mobility of American families and the increasing need for competencies in effecting smooth transitions and adapting to programs in education in other school systems nationally. Thus, there is a specific use for which standardized tests are intended. This use deals with broad norm bands which are intended to show general comparative trends in educational achievement when compared with the entire reference group.

Critics are numerous who attack the fact that standardized norm-referenced tests are not culture free and that they fail to validly test the non-reader or the poor reader. All written tests, regardless of their origin, have failed to adequately overcome cultural bias and reading

problems when used with large groups of pupils. However, measurement problems arise when faulty uses and uninformed, short-sighted interpretations are made with the scores of these pupils. Until more perfect instruments are devised, the standardized norm-referenced tests will continue to provide the best normative data for comparative studies in pupil achievement.

Attempts to measure the specific effects of an educational program on an individual must seek a more precise instrument of assessment which is custom designed to measure that particular program. Criterion-referenced tests offer a more specific and valid means of assessing individual performance and the accomplishment of program goals.

A criterion-referenced test is one that is deliberately constructed to yield data that are directly interpretable in terms of specified performance standards. This type of test is not designed to facilitate individual difference comparisons such as the relative standing of a pupil in a norm group or population. Such tests are not designed to enable one to speculate on a pupil's relative standing with respect to a variable such as reading ability. Instead, criterion-referenced tests are concerned with measuring an individual's performance relative to a specified domain of tests which includes both content and process.

Glaser was among the first to move in the promising direction of utilizing test items that were derived directly from the content of the behavior categories that were to be measured. He pointed out the distinction between the criterion-referenced tests which measured what a pupil could do and norm-referenced tests which compared a pupil's progress with that of others (Lipe and Jung, 1971 RER, October 1971).

Then, criterion-referenced tests offer distinct advantages when used to measure goal achievement and program effectiveness, to measure individual achievement in mastering the program objectives and to gain information about the placement of an individual in a continuum of specified skills within a program. Individualized instruction is particularly enhanced by criterion-referenced tests. Small increments in behavior can be detected by periodic small scale tests, enabling more frequent opportunities for incentive delivery. Further, when items are constructed to directly measure the degree of attainment of various behavioral goals of a program, an interesting marriage is achieved between constant behavioral observation and sporadic evaluation via norm-referenced achieve-

ment tests (Lipe and Jung, 1971, RER, p. 274).

As the feasibility of individualized instruction increases, knowledge of an individual learner's position in the group becomes less important than knowledge of the competencies that the individual does or does not possess. Hence, it is likely that educational assessment will require norm-referenced information in addition to criterion-referenced information.

The distinction between norm-referenced achievement tests and criterion-referenced tests can be found by (a) examining the purpose for which the test was constructed, (b) the manner in which it was constructed, (c) the specificity of the information yielded about the domain of instructionally relevant tasks, (d) the generalizability of test performance information to the domain, and (e) the use to be made of the test information.

Justification for the Adaptation of Norm-Referenced Test Items as a Criterion Reference Measure

Recognizing the professional skills of the psychometrists and statisticians found on the staffs of the recognized national test publishers, educators will be well advised to utilize the services of test consultants in devising criterion-referenced tests from those materials already purchased within a system. Few school systems or State Departments of Education have the staff, the financial resources, or the time to devote to the development of valid criterion referenced assessment materials.

In utilizing the professional services of testing consultants, the teachers, administrators or State Departments of Education should carefully examine the behavior categories that are to be specified in a test outline. A systematic plan must be devised to make sure that each try-out form of the test includes a representative sample of items in the behavior categories. In making this plan, the domains of items must be carefully examined and stratified to allow for a representative sampling. The terminal objectives or desired outcomes must be stated and the behavior which defines each point along the achievement continuum is carefully defined. The test items that test these behaviors are then selected with the guidance of the test consultants. The validation of the resulting test is established by the test publishers who provide further

services throughout the administration, scoring, interpretation and evaluation of the results.

Some areas of skill mastery are more readily adapted from standardized norm-referenced tests than others. The mathematics sub-tests are examples of this adaptability. It may be that a criterion-referenced test covering a wide domain is not likely to provide data that satisfactorily fulfills the basic purpose of such tests. It is suggested that for any given domain, a coordinated set of diagnostic sub-tests should be available each of which is made up of items that are homogeneous in the sense that they test performance on a specific behavior or on a cluster of behaviors that are taught as a unit.

There are numerous considerations involved in creating valid tests. Considerations involve such matters as: semantics, cultural bias, statistics, levels of difficulty, comprehensive coverage of domains, and even that of appealing format. Norm-referenced tests either have already met the publisher's criteria in such matters or the publisher can provide services to provide for these matters when a criterion-referenced test is adapted from items found in existing norm-referenced tests. Few professional educators could devote their time and limited resources to the production of criterion-referenced tests that could surpass the product that is developed jointly by educators and a testing publisher with their specialized resources.

Another factor favoring the adaption of criterion-referenced tests from norm-referenced tests is that there are already existing variety in the forms of tests at each level of difficulty. Most individualized instruction utilizes pre-tests, terminal tests done immediately after finishing a program, and post-testing done some time later. Norm-referenced tests are readily adaptable for assessing an individual's progress over a continuum of learning skills within each domain.

The most attractive factors involved in adapting existing norm-referenced test items to criterion-referenced tests are unquestionably the time saving and the money saving possibilities. Much has already been invested in extensive standardized testing programs in many school systems. Further use of these materials to satisfy assessment needs through careful adaptation of the materials to accomodate the intended purpose of the tests is logical and economical.

These factors then seem to support the practice of utilizing the materials at hand to develop criterion-referenced test items:

First - The need for criterion tests will increase as more systems adopt performance accountability models.

Second - Standardized norm-referenced tests are accepted and are, in fact, improving constantly.

Third - Educators can coordinate their instruction expertise with the test construction expertise of test publishers to produce an appropriate and satisfactory testing tool.

Fourth - The testing consultants can handle the technical considerations in test construction and scoring where few teachers feel secure in doing so.

Fifth - Appropriate, economical tests can be made available when needed without tedious and expensive delays.

Sixth - Test publishers have multiple forms of tests available to meet the needs for frequent criterion-referenced tests.

And finally - A school system has established a working relationship with a consulting service that can help teachers avoid problems in evaluation and will be available to advise as the programs progress and the final evaluation becomes necessary.

December 1972

BIBLIOGRAPHY

Glaser, R Instructional Technology and the Measurement of Learning Outcomes: Some Questions. American Psychologist, 1963, 18, 519-521.

Lipe, and S. M. Jung, Manipulating Incentives To Enhance School Learning. Review of Educational Research, 1971, 41:4, 249-280

Nitko, A. V., A Model for Criterion Referenced Tests Based on Use. 17p. (ED 049 318; MF and HC available from ERRS).

ED 078047

TM 002 862

Criterion-Referenced Test Development - A Contractual Arrangement Between the Public Schools of the District of Columbia and a Commercial Test Publisher

Mildred P. Cooper
Assistant Superintendent
for Research and Evaluation
Public Schools of the District of Columbia

FOREWORD

Everyone interested in educating youth is interested in developing as meaningful a program for the individual as is possible. To do this we must understand the needs of the individual student. In speaking of the instructional program needs of this student we must find a way to discover his strengths and his weaknesses in terms of the educational program which we offer him. Increasingly, school systems are turning to criterion-referenced testing to provide this necessary diagnostic information.

Ahead of other urban school systems and most school systems in the nation in recognizing the value and the relevancy of criterion-referenced testing, the Public Schools of the District of Columbia began in the fall of 1970 to develop such a testing program for its students. As the coordinator of that test development effort in which we utilized a commercial establishment, I have been requested to give a historical description and a chronological account of our experiences. The account is brief and yet is detailed.

CRITERION-REFERENCED TEST DEVELOPMENT IN READING AND MATHEMATICS

In 1970 the Board of Education of the District of Columbia engaged the services of Dr. Kenneth Clark to develop a plan to improve the academic achievement of the students in the District schools. There were many, many aspects of the plan, however, I will refer here only to those relating to the topic under discussion. The plan called for the establishment of minimum floors of achievement in reading and mathematics and also called for the administration of standardized achievement tests three times a year--in the fall, in January, and in the spring.

As you probably know from the reports in the news media, great turmoil occurred in the District of Columbia as a result of the adoption of the Clark Plan by the D. C. Board of Education. Teachers and other school staff had not been involved in the development of the "Design for Excellence", therefore, were especially antagonistic to certain features. Nevertheless, school staff did carry out the responsibilities that they had been assigned.

The reading specialists and the mathematics specialists in those subject field departments developed a series of sequential skills in reading and in mathematics at each grade level and designated those to be considered as minimum floors. These were then issued to the field by the Division of Instruction for use by teachers during the school year 1970-71.

In the meantime, the Pupil Appraisal Section of the Department of Pupil Personnel Services established the schedule for the three-time testing. The first administration of the standardized achievements tests was scheduled for late September, 1970. When the time came, many teachers refused to administer the tests and the Washington Teachers' Union demanded that the Superintendent of Schools, Dr. Hugh J. Scott, establish a Union-Board Testing Committee to come to some agreement on the testing policies and program. The Superintendent did establish that committee and designated me as chairman. After continuous meetings over a two-week period, a set of recommendations were proposed to the Superintendent. With almost no changes, the Superintendent approved these recommendations and they became the policies on testing for the Public Schools of the District of Columbia.

Within the testing policies, it was agreed that system-wide standardized achievement testing would not occur more than twice a year and that the school system would begin the development of instruments which would more relevantly measure the progress of the students in the District of Columbia. Conferences were held with measurement consultants and on the advice of Dr. Ralph Tyler a decision was made to use a different approach to the D. C. Public Schools' testing program. The direction recommended was the development of "mastery" or "criterion-referenced" tests related to the specific instructional objectives of the D. C. Public Schools.

After a further decision was made to limit the criterion-referenced test development to the areas of reading and mathematics, Superintendent Scott appointed me to coordinate the task. That was in December, 1970. Since a contract was in effect at that time with CTB/McGraw Hill, the logical approach was to negotiate the substitution of the development of the criterion-referenced tests for the already contracted materials and scoring services for that no longer desired third administration of the standardized achievement tests for that school year. It was then that I began my discussions with a representative of the company. After basic information had been communicated to the company, a meeting was held comprised of the heads of the Division of Instructional Services, Pupil Personnel Services, the Pupil Appraisal Section and the Departments of Research and Evaluation. Certain agreements on the plan of test development were reached. Subsequent to the meeting, the CTB representative and I had many conferences on procedures, funding, content, and other related items.

A proposal was developed and submitted by the California Test Bureau/McGraw Hill in January 1971. I forwarded copies of the proposal to the department heads who had been involved in the test development planning meeting. After their review, comments and appropriate changes, the proposal was submitted to the D. C. Public Schools Contracts Division for the development of the contract.

With a contract document which spelled out most of the terms, I met with the District of Columbia Government Negotiated Services Chief, a Vice-President of CTB/McGraw Hill, the D. C. Public Schools Contracts Specialist and the designated Project Director for CTB/McGraw Hill. The purpose of this meeting was to "hammer" out a clause which would permit

the D. C. Public Schools to recover the contractual costs of the test development through a reimbursement or discount on the purchase of the test instruments over a period of time. Agreement was reached and by the end of 1973 the total contractual development costs will have been recovered by the D. C. Public Schools.

My review so far has been an account of the events and the procedures leading up to the development of the contractual agreement with the test publisher. Although this phase took a great deal of time both on my part and that of other school staff, it was small in comparison to the amount of time and effort required in the implementation of the contract. The greatest problem that I encountered in the criterion-referenced test development project was that no resources were allocated for the test development other than the limited funds for the contractor. The work done by all of us from the school system was in addition to our regular responsibilities and assignments. Test development is much too complex a task not to have individuals provided with ample work time to carry out the task. Most people outside the field of measurement and evaluation do not understand the complexities of test development.

The initial step in the implementation of the contract was the development of a work flow chart by the CTB Project Director and me. A copy of this chart appears on the next page. Looking at the chart and with the amplifications which follow, the chronology, the problems and the actions of the project can be traced:

2.0 Review of D. C. objectives -

2.1 Identify D. C. objectives -

The District of Columbia objectives in reading and mathematics had to be thoroughly reviewed to determine the congruence between objectives and the instructional program then operating.

a) The first task was to get written verification and confirmation of objectives.

b) The revision of mathematics objectives in 1971 caused delays in the schedule.

c) The committees of reading and mathematics specialists then reviewed objectives to be sure all were in line with instructional program goals and designs.

d) As the coordinator I tried unsuccessfully to get funding for work during the summer of 1971, therefore, reading specialists worked whenever they could fit it into their schedules.

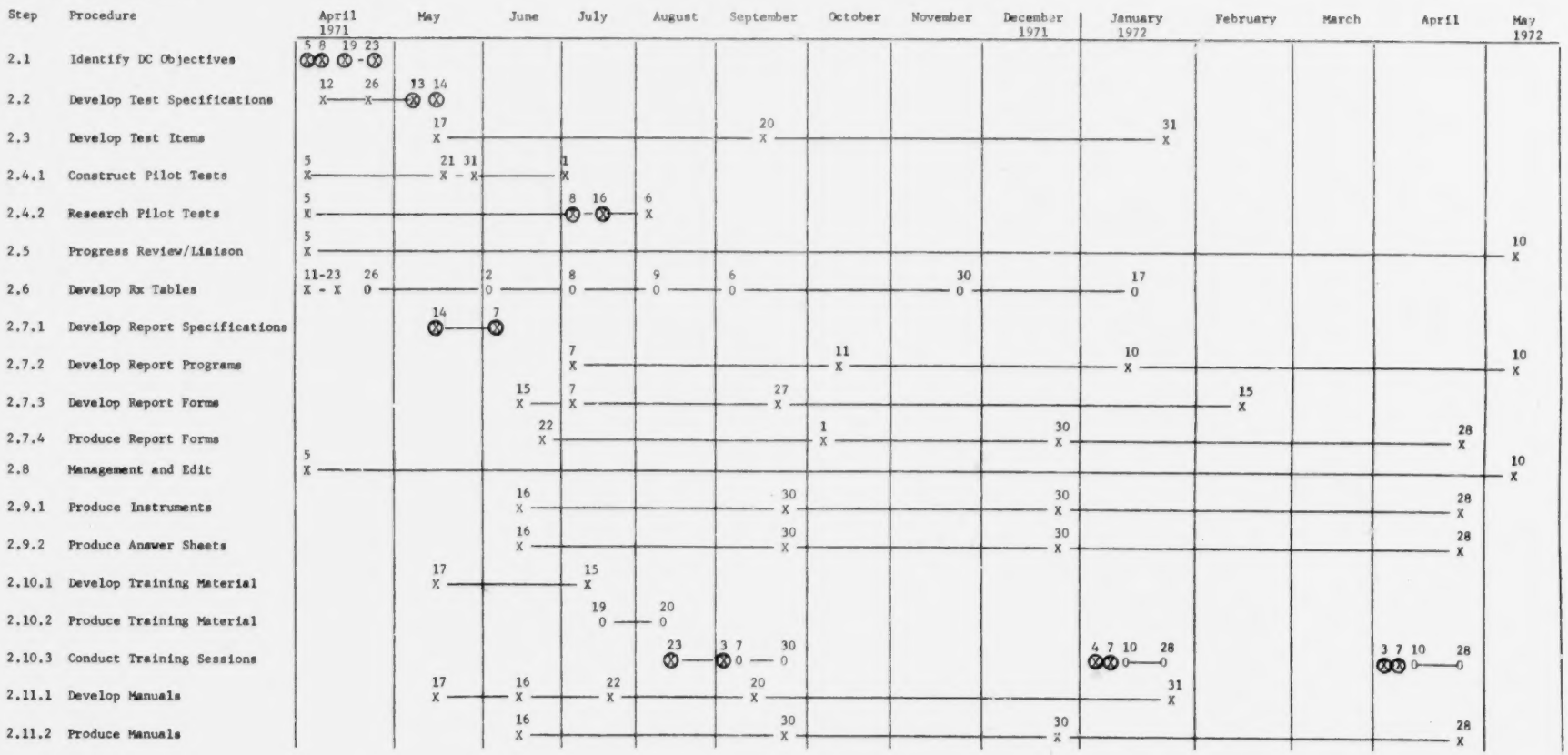
2.2 Develop test specifications -

In all test instrument development it is necessary to identify

Public Schools of the District of Columbia
Departments of Research and Evaluation

DEVELOPMENT OF CUSTOMIZED DC-PMT AND DC-PRT

X -- CTB Responsibility
O -- DC Responsibility
⊗ -- Joint Responsibility



certain factors which influence the format and content of the instrument. In this test instrument development the following were considered:

a) Level of tests

b) Number of objectives to be measured

c) Overlap between levels

d) Number of items per level

e) Types of response

f) Problems of non-graded and other class structure

The above items were discussed at meetings with testing staff and instructional personnel.

2.3 Develop test items -

Test items had to be written to measure the objectives from 2.1 and to meet the specifications of 2.2.

a) Decisions on factors in 2.2 had to be observed.

b) Test items used were those developed by teachers, test specialists and subject field specialists. Items included were validated items.

c) Review was made by D. C. Public School reading and mathematics specialists.

d) In reading, selections to be used were scrutinized for appropriateness.

e) Where the specialists felt changes were needed in items or selections, such suggestions were forwarded to CTB. In all instances those changes were incorporated.

2.4.1 Construct pilot instruments -

Because of problems of vocabulary level and approach at the early childhood years, it was mutually agreed that a pilot study was mandatory if an appropriate instrument were to be developed.

a) Tests were for grades Kindergarten, 1, 2 and 3.

b) New test items were developed.

c) Conferences were held with D. C. school personnel relative to the pilot instruments.

2.4.2 Research pilot test -

The pilot testing was done in spring 1972 in a sample of schools across the D. C. Public School system. This pilot was necessary in order to observe student reactions to the response modes under study.

a) Levels K-3 were included in the pilot project.

b) A carefully designed sample of schools and students was developed.

c) The pilot study was coordinated by a staff member of the Departments of Research and Evaluation. Conferences with principals and teachers were required prior to the conduct of the study.

d) The pilot study required 12 professional staff members from the

Departments of Research and Evaluation and the Pupil Appraisal Unit of the D. C. Public Schools and the California Test Bureau.

e) The processing of the data was handled by CTB after it had been collected in the D. C. Schools.

f) The new instruments based on the findings of the pilot project followed the regular review and revision procedures of instruments for grades 4-9 and will be ready for administration in the spring of 1973.

2.5 Progress review and project liaison -

This activity began with the discussion of the original proposal. This continuous communication was and is very time consuming, but is of extreme importance. Contacts by mail, telephone, and on-site consultations were required.

Progress review and project liaison are extremely important for the success of the entire project is dependent upon clear communication, evaluation of work and appropriateness for the local school system.

a) A coordinator must be, and was, designated.

b) Communication with CTB Project Director was a continual weekly, and usually more frequent,event during the course of the contractual agreement.

2.6 Develop instructional prescription tables -

The ultimate value of the prescriptive criterion-referenced test lies in its support for teacher and student for the individualization of instruction and improvement of the learning process. Such value rests squarely upon absolute congruence between the objectives and the instructional resources available for teacher and student.

Instructional resources (such as texts and reference material, cited by title and page number), must be identified for each objective.

a) This necessary, time-consuming task which relates curriculum materials to the test was undertaken by D. C. school personnel as a part of the project.

b) As the coordinator I tried to get extra funding during the spring and summer 1971 for the task; none was made available, therefore, Dr. James T. Guines, Associate Superintendent, Division of Instructional Services, detailed 12 reading specialists from their regular assignments to the task for 2 weeks in September 1971.

c) The task of getting all curriculum materials keyed is expensive in terms of staff time but is a mandatory part of the project so not only staff of the Division of Instruction but key staff in Departments of Research and Evaluation and staff in Department of Automated Information Services worked on this.

d) Several meetings to train individuals in the process were held by the CTB coordinator and the D. C. schools project coordinator.

e) Instructional format and procedures for keying prepared by CTB did not fit D. C. circumstances, therefore, adjustments were made and the revised formats utilized.

f) The keying of a limited number of curriculum materials at each level for grades 4-9 in reading and mathematics was accomplished by D. C. school personnel and CTB staff before the test administration in the fall of 1972.

2.7.1 Develop report specifications -

The report specifications were an outgrowth of item 2.2. The data treatment, data to be reported to whom and the format were agreed upon by the D. C. and CTB coordinators and appropriate D. C. school staff. Format changes have been and will continue to be made as a result of the utilization of the instruments.

2.7.2 Develop report programs -

The data processing staff of the California Test Bureau designed and wrote the necessary computer programs for the above.

2.7.3 Develop report forms -

The California Test Bureau's data processing and manufacturing departments worked with the CTB coordinator in the development of the necessary reporting forms.

2.7.4 Produce report forms -

The manufacturing department of the California Test Bureau contracted out the job and the report forms were printed.

2.8 Management and editing materials -

The California Test Bureau staff assumed all responsibility for correctness of all copy for syntax, context and format.

2.9 Produce instruments -

The production of the test booklets PMT - DC and PRT - DC was in the hands of the CTB Department of Manufacturing and Word Processing. The booklets for levels D through H in reading (grades 4-9) and for levels E through H in mathematics (grades 5-9) were produced and ready for the fall 1972 testing program of the D. C. Public Schools.

Levels A through C in reading and levels A through D in mathematics are being completed and will be ready for test administration in the spring of 1973.

2.10.1 Develop training program for D. C. staff:

The CTB staff and staff of the D. C. Public Schools' Pupil Appraisal Section developed training materials on the use of the prescriptive test results for the individualization of instruction. In addition to the materials, the plan for the training sessions was also developed.

2.10.2 Produce training materials -

The actual production of the training materials was done by CTB and D. C. Public Schools and provided to classroom teachers and other staff.

2.10.3 Conduct training sessions -

Staff of the California Test Bureau and the staff of the Pupil Appraisal Section of the D. C. Public Schools conducted training on

the administration of and use of the prescriptive test instruments with the testing chairmen of each elementary and junior high school of the D. C. Public School system. These testing chairmen with assistance from Pupil Appraisal decentralized staff then subsequently conducted training sessions for the classroom teachers in his or her building prior to the spring 1972 grade 4 and 6 reading testing program and the fall 1972 city-wide testing at grade levels 4-9 in reading and grade levels 5-9 in mathematics.

2.11.1 Develop manuals -
and
2.11.2 Produce manuals -

The California Test Bureau staff developed and produced manuals for teachers for each level of the tests. These manuals contain specific directions to teachers for preparations for students to take the test, actual administration of the test and preparing test answer sheets and booklets for scoring.

Manuals have been used for grades 4-9 in reading and grades 5-9 in mathematics. Manuals for grades 1-3 in reading and mathematics and in grade 4 in mathematics will be available when the PRT-DC and PMT-DC are used at those levels.

These were the steps and actions in the criterion-referenced test development for reading and mathematics grades 1-9 in the Public Schools of the District of Columbia. A more complete description will gladly be given upon request.

* * * * *

What problems were encountered in the District of Columbia in the project to develop criterion-referenced tests? The major ones were:

(A) There existed a severe lack of resources to do the job.

(B) Criterion-referenced test development was not given the level of priority in and by the school system commensurate with the task.

(C) It was difficult for a long-term project of the magnitude of criterion-referenced test development to comply with a carefully designed work flow chart when programs, priorities and objectives of the school system were going through a series of changes. In this kind of project many related tasks such as printing and data processing must be scheduled far in advance.

What were the outstanding positive elements in the developmental project? Among them were:

(A) The planning and working together of a great many operating departments of the school system to accomplish the criterion-referenced test development task.

1) The exceptional cooperation and hard work of the Supervising Directors of Reading and Mathematics and their staffs of reading and mathematics specialists and of classroom teachers in the developmental and review procedures and in the preparation of prescriptive data.

2) The keen interest and the willingness of the Directors and staff members in the Department of Automated Information Systems and the Pupil Appraisal Section to perform required tasks in the developmental process of the criterion-referenced tests in addition to their assigned work-load.

(B) The opportunity provided through this project to various school personnel including teachers to gain first-hand experience in test development for system-wide use; subsequently, the use of these tests and test results will have a major impact on the instructional program at the classroom level as well as on the overall instructional plan.

(C) The positive reactions of the subject field specialists and most teachers as evidenced by written comments and oral statements which indicated that:

1) They felt that at last tests were available that were directly related to the instructional objectives of the local school system.

2) They felt that the tests had more appeal to students than the ones previously used.

(D) The results of the fall administration of the tests gave to teachers diagnostic prescriptive information on each individual child that took the test, thus providing valuable instructional assistance.

I would make very strongly the following recommendations to school systems planning to develop criterion-referenced tests:

(A) The school administration and the Board of Education must be committed to support the concept of criterion-referenced testing.

(B) Adequate resources must be allocated to support the entire project.

(C) A full-time coordinator or director with no other assignment should be appointed before any planning steps are undertaken.

(D) The school system must have clearly defined instructional objectives. These objectives must be stated in measurable terms.

(E) There must be a clear understanding by the school system of the purposes of various types of measurement and the appropriate use of various measures.

In closing I would like to say that the only complaints I have had so far from teachers and principals in the D. C. school system refer (1) to the fact that not all books and materials are keyed and (2) that the 4th grade PMI test developed by CTB for national use was not appropriate for the District of Columbia. (The D. C. edition was not developed in time for the fall 1972 test administration program).

As the school system's coordinator I might summarize my reaction to the criterion-referenced test development project in this way: I feel that in addition to the value of the criterion-referenced tests to the instructional staff and program, as cited above, the comments from pupils have made the project completely worthwhile. They say, "we like these tests much better than the ones we used to take. They are much more interesting."

December 1972

The Development of Criterion-Referenced Tests in Reading

Richard M. Petre
Consultant in Reading
Maryland State Department of Education

The Maryland State Department of Education established reading as a priority approximately three years ago. One of the components of the plan was to implement the following as outlined by the Department's executive staff:

- To ensure that each student and adult possess the basic skills necessary to become an effective citizen.

- By 1977, 85 percent of all Maryland students will be able to use reading as a communicative skill as determined by appropriate criterion-referenced measures.

A committee representing teacher-training institutions, local school systems, and the State Department worked with Dr. Roger Farr of Indiana University, to determine the best approach to accomplish these objectives.

To define the problem specifically, the committee introduced the following four objectives:

- By 1977, all students enrolled in the public schools, excluding permanent care institutional cases, who have completed an elementary school program will be able to use independently the communication purposes of reading outlined in Table I, Parts I and II.C.1.

ED 078048

TM 002 863

- By 1977, 85 percent of all students completing a public school elementary program will be able to go beyond objective 2.2a and will be able to demonstrate a variety of performances of the process of reading outlined in Table I, Parts I and II.C.1. as determined by appropriate criterion-referenced measurements.

- By 1977, all students enrolled in the public schools, excluding permanent care institutional cases, who are 15 years old will be able to use independently the comcumication process of reading outlined in Table I, Parts I and II.D.1. as determined by appropriate criterion-referenced measurements.

- By 1977, 85 percent of all students completing a secondary school program will be able to go beyond objective 2.3a and will be able to demonstrate a variety of performances of the communication processes of reading outlined in Table I, Parts I and II.D.2. as determined by appropriate criterion-referenced measurements.

The standard "85 percent of all students" was changed to a 100 percent level because some reading behaviors are needed by all people, with most students going beyond the basic level. The objectives also specified due dates for mastery at the following levels: 12 year old, 15 year old, or ready for high school graduation.

In the firm belief that survival in society is the paramount reason for teaching reading in school, the committee selected functional reading as the priority emphasis. Five basic functional purposes for reading instruction were agreed upon: (1) following directions; (2) locating references; (3) personal development; 94) gaining information; (5) using forms. Table I one the following pages represents the combined efforts of the committee and other groups - in Maryland who have agreed that the contents represents those survival reading tasks which are essential. This content source outline was used by SEE, Inc., of Bloomington, Indiana as the basis for writing the criterion-referenced tests to be used in Maryland.

TABLE I

Content Source Outline For Maryland's Criterion-Referenced Tests

Objective	Sub-Objective		Source	Age Level
I. Following Directions	A. Basic vocabulary	1.	EDL Revised Core Vocabulary Grades 1-8, 9-13	12 15 18
		2.	Maryland driver's manual	
	B. Sequential order	1.	Cooking recipe	12 15 18
		2.	Sewing pattern directions	15 18
		3.	Emergency directions (in case of fire)	12 15 18
		4.	Game directions (monopoly, e.g.)	12 15 18
		5.	Build-it-yourself model directions	12 15 18
		6.	Do-it-yourself kit directions	12 15 18
		7.	Textbook directions	12 15 18
		8.	Test directions	12 15 18
		9.	Boy Scout manual - activity directions	12 15 18
		10.	Cleaning solution directions	15 18
		11.	Voting directions	18
		12.	Child care directions (bathing, e.g.)	15 18
		13.	Directions for running household appliances	15 18
		14.	Directions for first aid (burn treatment, snake bite, artificial respiration	12 15 18
		15.	Medicinal directions (dosage, use of thermometer)	15 18
		16.	Telephone directions	12 15 18
	C. Cautions/Labels Warnings	1.	Warning signs (danger, high voltage, polluted water)	12 15 18
		2.	Medicine labels	12 15 18
		3.	Warning labels (poison, caustic, harmful if swallowed)	12 15 18
		4.	Food labels	12 15 18
		5.	Warnings related to care of children and possessions	15 18
	D. Finding location	1.	Street signs	12 15 18
		2.	Maps - local, state, country, subway	12 15 18
		3.	Bus schedules and maps	12 15 18
		4.	Work schedule	18

Table I (continued)

Objective	Sub-Objective		Source	Age Level
II. Locating References	A. Single volume	1.	Table of contents	12 15 18
		2.	Index	12 15 18
		3.	Footnotes - bibliography	12 15 18
		4.	Glossary	12 15 18
		5.	Headings, subheadings	12 15 18
		6.	Catalogue	12 15 18
		7.	Phone book and Yellow pages	12 15 18
		8.	Appendix	15 18
		9.	Almanac	18
		10.	T. V. Guide	12 15 18
		11.	Newspaper articles	15 18
		12.	Dictionary	12 15 18
	B. Multi-volume	1.	Encyclopedia	12 15 18
		2.	Magazine index	15 18
		3.	Reader's Guide	18
	C. Multi-resource	1.	Card catalogue cards	15 18
		2.	Reference books and materials	18
III. Personal Development				
IV. Gaining Information	A. School	1.	Vocabulary lists	12 15 18
		2.	Textbook material - content areas	12 15 18
	B. Job	1.	Training manuals	18
		2.	Safety requirements	15 18
		3.	Want ads	18
	C. Society	1.	Public announcement	15 18
		2.	Legal documents	18
		3.	Emergency announcements	15 18
		4.	Political material	18

Table I (continued)

Objective	Sub-Objective		Source	Age Level
IV. Gaining Information (continued)	D. Home	1.	Newspapers	15 18
		2.	Consumer's Guide	18
		3.	Ads	15 18
		4.	Service bills	18
		5.	Government pamphlets	18
		6.	Day care information	18
V. Forms	A. Personal Information	1.	School forms	12 15 18
		2.	Driver's license or learner's permit application	18
		3.	Work permit application	18
		4.	W-2 form	18
		5.	Armed forces forms	18
		6.	Social Security forms	18
		7.	Insurance forms	18
		8.	Medical forms	18
		9.	Welfare application	18
		10.	Job application	18
		11.	Cereal box top	12
	B. Financial	1.	Subscriptions	15 18
		2.	Sales slip	12 15 18
		3.	Mail order purchase slip	15 18
		4.	Credit card	18
		5.	Notes, loans	18
		6.	Bank statement	18
		7.	Rent agreement	18
		8.	Long-term purchase agreement	18
		9.	Income tax form	18

We believe that Table I is one of the most comprehensive lists of survival tasks based on societal demands in the literature on reading. We also feel if Maryland students can perform these reading tasks they have the basic reading knowledge needed to function in society as well as those basic skills which will enable them to handle new reading situations throughout their lifetime.

SEE, Inc. prepared for us three criterion-referenced tests: basic and advanced items for mastery by 12 year-olds, basic items for mastery by 15 year-olds, and advanced items for mastery before high school graduation. Samples from these tests are given below:

- -

SAMPLE ITEMS ON THE CRITERION-REFERENCED TEST:

Directions:

Read each of the following questions carefully and mark the answer that describes you best.

1. How much time do you spend reading for fun during vacations?

   a. None
   b. One to three hours a week
   c. Three to six hours a week
   d. More than six hours a week

2. How do you feel about reading as a spare time activity?

   a. I enjoy it
   b. I can take it or leave it
   c. I'd rather do something else
   d. I don't like it at all

This newspaper index will help you answer questions 3 and 4.

THE INSIDE STORY

Fair, Colder
Slightly Warmer
On Thursday

(More Weather on Page A12)

MICHIGAN CLINCHES 1ST PLACE in Big 10 Conference with 79-67 win over Northwestern. This and other sports on Page B4.
INSURANCE IN WORKS for city employees. Read this C-T editorial today on Page A10.

Ann Landers	A9	Movies	B2
Bridge Column	B13	Obituaries	B14
Business News	B3	Question Girl	A5
Classifieds	B10-11	Sports	B3-7
Comics	B12-13	State News	B9
Crossword	B13	Statistics	A12
Editorials	A10-11	Television	B12
Family Living	A8-9		

3. Which of the following would be found on page B12?

   a. Baseball scores
   b. Weather information
   c. The time of a T. V. Program
   d. A crossword puzzle

4. What section would you look at to find the cartoons?

   a. Comics
   b. Movies
   c. Sports
   d. Television

The grocery ad in the box is for items 5 and 6.

SAVE 10¢ PER LB.
LESSER AMOUNTS
75¢ PER LB.

Fish Steaks	2 lb. $1.19
Deep Sea Favorite Frozen Turbot 3 lb. or more	lb. 69¢
Select Sliced Beef Liver	lb. 59¢
Honeysuckle 12-14 lb. avg. Young Turkey	lb. 49¢
Skinless Wieners	12 oz. pkg. 59¢
Large Bologna By the Piece	lb. 59¢

Mark A for True and B for False.

5. Turbot is a sea food.

   a. True
   b. False

6. The cheapest meat per pound in the market is hamburger.

   a. True
   b. False

Mary just bought the groceries. Use the grocery tape in the box to the left to answer each question below.

.	00.39	GR
.	00.43	GR
.	00.72	GR
.	00.22	GR
.	00.16	GR
.	00.17	GR
.	00.17	GR
.	00.16	GR
.	00.33	GR
.	03.65	MT
.	05.88	MT
.	00.25	TX
.	12.53	sb t1-

.	00.46	GR
.	00.36	GR
.	00.02	TX
.	13.38	t1-
ca		
td	20.00	--
.	06.62	CG--

THANK YOU
-4-

7. How much money did Mary give the cashier?

   a. $12.53 c. $20.00
   b. $13.38 d. $ 6.62

8. How much change did Mary get?

   a. $12.53
   b. $13.38
   c. $20.00
   d. $ 6.62

Use the oven operating instructions in the box to answer questions 9 and 10.

Mrs. Jones has just bought a new electric range. Here are the operating instructions for the oven.

1. The reading on the oven thermostat dial shows BAKE area from 150^{o} to 500^{o} and BROIL area from 375^{o} to "Broil."

2. Baking

   Turn dial to desired temperature. If the dial is set above 300^{o} both broil and bake elements stay on until desired temperature is reached when the broil element goes off. You will know when the desired termperature is reached as the indicator light will go off.

Mark the letter on your answer sheet that is the best answer to each question.

9. What is the temperature range for BAKING?

   a. 100^{o} to 200^{o}
   b. 150^{o} to 300^{o}
   c. 300^{o} to 375^{o}
   d. 150^{o} to 500^{o}

10. How does this electric range preheat the oven quickly?

   a. Both broil and bake elements stay on
   b. The broil element stays on
   c. Both broil and bake elements stay off
   d. The bake element stays on

In the box on the left is a list of companies who are looking for employees. To answer questions 11 - 14 mark the letter on your answer sheet which corresponds to the job for which that person is best qualified.

A. OPENING for experienced linotype operator and a composition floor or lock-up man. All inquiries kept confidential. Contact Mr. Bobby Hall, Midland Press, Inc., Spencer, Maryland.

B. EXPERIENCED COOK needed. Lunch and dinner hours. Good pay, good working conditions. References required. Apply in person. MIKE'S CAFE, 217 N. Walnut.

C. IN DESPERATE need of lead guitar player for rock group. Please phone 336-5166 or 339 8317.

D. RN or LPN full time or part time 7-3:30 and 3-11:30 shift available. Pay commensurate with experience. 339-1657, after 5 p.m. 336-5570

EVENING COOK WANTED. Hours 12 noon to 8:30 p.m. Will train. Good starting wage and benefits. Apply in person between 2 p.m. and 4 p.m. ABC CAFETERIA, College Mall.

11. ___ Jane MacDonald
Registered Nurse

12. ___ Mr. J. Fish
Experienced

13. ___ Bobbie Haven
Guitarist

14. ___ D. Y. Nelson
Printer

Our first statewide sampling of the test items was conducted in May, 1972. This was the second trial for the test items. The following results were completed:

Basic Test: Includes those items we want 100% of the students to achieve. (Successful performance of the items was 80% or above)

Age	No. of Students	% of Students Successful
12 year old	470	65
15 year old	416	84
High school graduate	301	85

Advanced test: Includes those items we want most students to also achieve. (Successful performance of the items was 80% or above)

Age	No. of Students	% of Students Successful
12 year old	512	45
15 year old	416	65
High school	209	72

As you know, we in Maryland are exploring a different approach to reading. We have chosen to emphasize survival reading. Accountability for results is based on the following:

(1) Declaring specific goals and behaviors for three age groups.

(2) Testing observable performance on items needed for survival reading instead of those skills usually measured on standardized tests.

(3) Planning instructional decisions based on the results of testing described above.

Currently, we are planning to test in each local school system. A second form of the criterion-referenced measurements will be constructed. Guidelines will be written to help local school systems implement functional reading as a part of their already ongoing reading programs.

Our State priority is based on the belief that a reader is one who not only can read but does read. Logically, then, the place to start is with functional reading needs. Thus, our effort and this working report.

December 1972

ED 078046

TM 002 864

The Development of a Criterion-Referenced Test of Mathematics

Fanny Freytes
Director of Program Evaluation in Mathematics
Department of Education, Puerto Rico

INTRODUCTION

During the last few years the Puerto Rico State Department of Education has made great efforts to provide the educational system of Puerto Rico with a series of adequate standardized tests which are essential for assessment and evaluation purposes. These efforts have included the development of standardized achievement tests for all levels in almost all of the curriculum areas. However, in Puerto Rico as well as the United States with the advent of increased federal aid to education and the subsequent emphasis in the areas of educational planning, individualized instruction and accountability there developed an awareness and a concern for the need of other types of instruments, namely criterion-referenced tests. Criterion-referenced measures have been considered particularly desirable in areas where diagnostic information is needed, such as placement of individuals in programs of instruction, formative evaluation of educational programs, and in evaluative assessment of individual or group achievement.

A first attempt toward the development of a criterion-referenced test has been undertaken as a joint project of the Division of Evaluation

and the Mathematics Program in the Department of Education. The project is concurrently developed with an accountability project in the area of mathematics at the seventh grade. As expected in an accountability project, information regarding specific skills and knowledge development is required in addition to the normative base for comparisons and interpretations.

Accomplishments

As an initial step in the development of this instrument, a careful review of the literature on the subject of criterion-referenced testing was done. The term "criterion-referenced" appears to have been introduced by Robert Glaser (1963) in a paper in which he distinguishes "criterion-referenced" from "norm-referenced" testing. In the latter, an individual's test performance is interpreted with respect to the performance of other individuals who belong to some specified population. In contrast, the interpretation of an individual's performance on a criterion-referenced test is a behavioral statement that is made without reference to the performance of other individuals.

Although considerable amount of attention has been given to the subject of criterion-referenced measures, there are very few guides available to the constructor of this type of instrument and in some cases the prevailing ideas are of a contradictory nature. In 1970, Dr. Stephen Klein from the UCLA Center, in his analysis of the relative efficiency of tests as vehicles for providing information for decisions about students and the educational programs they receive, suggested a four-step procedure which intends to combine the better components of the norm and criterion referenced test approaches. The essential characteristic of this approach is that it includes the concepts of item difficulty and normative score reporting in the development and interpretation of criterion based measures. This approach includes the following steps: 1) Specification of objectives, 2) Developing test items for each objective, 3) Developing test items to measure related objectives, 4) Providing score and score interpretation for each objective. Each of these steps will be taken up in turn.

The purpose of theAccountability Project which the mathematics test is intended to serve, as stated in the "Tentative Draft of Accountability Model for Mathematics Achievement" (See Appendix A) require information on the progress attained by individual pupils, by classroom, by school,

by level, by zone, by district, by region, by region-levels and by region-zones as compared with the criteria for normal achievement. It was with these purposes in mind that it was decided to follow Dr. Klein's suggested approach in the development of this instrument.

Clarification of Objectives

The goals for the teaching of mathematics in the elementary grades of the school system of Puerto Rico are clearly specified by the Department of Education and apply to all children at the elementary level. The mathematics test to be developed will be based upon these objectives and will attempt to measure to the greatest degree possible the specific skills, concepts, abilities and knowledges expected to be acquired by the end of the sixth grade.

The test will be divided into sub-parts according to the grouping of objectives by content areas. These are: Basic Concepts of Set Theory, Numeration Systems, Operations, Geometry, Measurement and Graphs.

Development of Test Items for Each Objective

Approximately 50 items have already been developed by program specialists who have had previous training in item writing and test construction techniques and were further reviewed by personnel from the Division of Evaluation.

One hundred or more items are being developed at present. In the development of these items special care is being taken to have a good representative sample of the total population of items that might be used to measure the objectives considering both the range of formats and the range of item difficulty.

Development of Test Items to Measure Related Objectives

Learning in mathematics goes through sequential stages: the understanding of one concept, the acquisition of one skill, is basic to another. If the foundation is poorly made, the structure as a whole will be weak and inadequate. Thus, it is very important to assess performance both on objectives that are either easier or more difficult to master rather than just the ones of major interest. This point is being carefully considered in the development of items previously described. As indicated by Dr. Klein in his paper on this subject, the reasons for measuring these kinds of related objectives are that they (a) provide information about the

unanticipated outcomes of educational programs (b) indicate how close a program (or student) came to meeting or surpassing the objectives (c) show the level at which subsequent educational treatments should be pitched.

Providing a Score and Score Interpretation for Each Objective

As the interest of this project is specifically directed toward attainment of specific objectives, the results on the test will be analyzed in terms of these objectives. The information provided by the test should reflect both criterion and norm-referenced performance on the items designed to measure the objectives. As previously indicated the mathematics test being developed will encompass the six global objectives listed for the sixth grade. The idea is to have subscores for each part of the test designed to measure each objective.

The statistical data relating to analysis of test results in terms of objectives of the mathematics program will be grouped into two major modalities. First, of all statistics on achievement in accordance with total number of items assigned by parts and subparts will be made. Under the criterion of score norms, statistics will be compiled (1) on the median, first and third quartile scores by subparts (2) on the percentage of items correctly answered in each sub-part (3) on an item-by-item tabulation of percentage of errors, and (4) percentage level of achievement in accordance with score norms.

Statistical data will also be presented tending to indicate the degree of attainment of objectives in terms of "expected" achievement. The criterion of expected level of achievement will be determined by program specialists at the State Department level and by experienced teachers at the local level. They will make their judgments on the basis of attainment they expect students to make at the end of the school year on each item of the test: 100 percent, 75 percent, 50 percent or less than 50 percent. The statistical compilations will include (1) the percent level of expected achievement item by item as determined by teachers and program specialists (2) a comparison between median obtained scores by sub-parts with the criterion of expected median scores (3) a comparison between the expected and attained percentage of level of achievement.

We should then be able to determine with some degree of accuracy how closely sixth grade children are achieving goals in mathematics in terms of norm scores and in terms of expected scores and will be able to use this information as the data base for the assessment of progress attained by students in the seventh grade, which is the grade to be considered in the Accountability Project previously mentioned.

December 1972

REFERENCES

Brazziel, William F. Criterion-Referenced Tests - Today's Education. NEA Journal, Nov. 1972. pp. 52-53.

Davis, Frederick B. Criterion-Referenced Measurement. T M Reports, 1971, Number 12

Flanagan, John C. Discussion Educational and Psychological Measurement, 1962, Vol. XXII, No. 1. pp. 35-39

Fremer, John. Criterion-Referencing. CSEA NEWSLETTER, October 1972.

Fremer, John. Criterion-Referenced Interpretations of Survey Achievement Tests, Test Development Memorandum. January 1972.

Hambelton, Ronald K and Novick, Merlvin R. Towards an Integration of Theory and Method for Criterion-Referenced Tests. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago. 1972.

Klein, Stephen. Evaluation Tests in Terms of the Information They Provide. UCIA Evaluation Comment, 1970, Vol. 2, No. 2.

Popham, W. James. Criterion-referenced Measurement. Englewood Cliffs, N. J.: Educational Technology Publications, 1971.

Appendix A

TENTATIVE DRAFT OF ACCOUNTABILITY MODEL FOR MATHEMATICS ACHIEVEMENT
1972

I GOALS	II. CRITERIA FOR NORMAL ACHIEVEMENT	III. PROGRESS ATTAINED	IV. INTERPRETATION OF RESULTS (Determination of needs)
	Minimum requirements or progress norms for the types of skills to be included in the program will be outlined. (The criteria for normal achievement, as well as progress norms, will be determined by the program concerned, in this area, mathematics in the seventh grade.	Compared with the criteria for normal achievement, what progress has each pupil attained? 1. by grades 2. by classrooms 3. by schools 4. by levels 5. by zones 6. by districts 7. by regions 8. by regions-levels 9. by regions-zones This will be ascertained through test prepared by the Department of Education or outside the Department expressly for this purpose. The type of test to be prepared and to be used will depend on the kind of skill being tested. Other evaluation measures that could be used are student records, interviews, etc.	Once the pupil's answers have been processed and analyzed, or once the "grades" obtained by the pupils are ready, different divisions of the system will interpret the results in terms of the various factors that may have influenced them. These interpretations will be in addition to those made by the Division of Evaluation. The divisions to be consulted will be units within the following: 1. program directors 2. regions 3. districts 4. schools 5. classrooms 6. professionals outside the system Influencing factors will be compared in at least three levels of achievement: deficient, normal and superior. In order for the interpretation of results to include a consideration of prevailing circumstances for each pupil, information on the following factors will be collected and studied: 1. clientele (pupil, home, community) 2. curriculum 3. physical facilities and instructional materials 4. teachers 5. supplementary services 6. teaching procedures (organization, methodology, measurements, evaluation, etc.)

USE OF RESULTS	VI. PROGRAM TO BE DEVELOPED IN TERMS OF NEEDS	VII. SYSTEMATIC EVALUATION	VIII. DEFINITION OF RESPONSIBILITY OF EVERY PERSON INVOLVED THROUGHOUT THE WHOLE PROCESS
Results will be used for:	In-service training to: 1. Improve the capability of persons with specific evaluation responsibilities at the different levels - a. significance and use of evaluation in educational systems b. use and preparation of different types of tests 1. standardized 2. criterion reference 3. diagnostic 4. achievement 5. those prepared by the teacher, etc. 2. Improve the ability of teachers to use different tests and their results, as well as to prepare tests for regular use. 3. Strengthen ability of program directors and curriculum technicians, etc. in conciliating and correlating test results with curriculum guides. (The overall training will include the fact that evaluation operates within a system). If the training were limited to persons specialized in evaluation, it could easily develop into so sophisticated an evaluation program that the system could not use it or area refuse to use it.	Along with the implementation of the action plan and the training program a systematic evaluation will be conducted. This will stress different types of evaluation, such as: 1. process evaluation 2. progress evaluation 3. product evaluation 4. formative evaluation 5. summative evaluation	As may be seen, this developmental process is a joint labor involving many areas of the educational system. For its success, it would be necessary to outline precisely the specific responsibilities of each area within the overall process. Some areas concerned with the process are: 1. Secretary of Education Certain specific responsibilities here would be: a. convene and preside over the initial meeting on the process in question b. assign the responsibilities to each area c. determine the kind of public relations work to be performed d. authorize use of the necessary resources 2. Area of instructional programs 3. Area of planning and development 4. Educational regions 5. School districts 6. School principals 7. Teachers 8. Students 9. Parents 10. Others